

REGRESSIOONANALÜÜS SORDIVÕRDLUSKATSETES

E. Lauk

Väetiste optimaalsete normide kindlakstegemisel on soovitatav põldkatse metoodika üles ehitada selliselt, et uurimistöö tulemusi oleks võimalik üldistada regressioonanalüüsi abil. Seejuures võib piirduda põldkatses ainult ühe kordusega, suurendades samal ajal tunduvalt variantide arvu, kuna regressioonanalüüs võimaldab välja arvutada katsevea ka ühe korduse olemasolul (Lauk, 1994; 1995).

Sarnast metoodikat kasutatakse Rootsis taimekaitsevahendite optimaalsete normide kindlaksmääramisel ning selleks on välja töötatud sellist tüüpi põldkatsete läbiviimiseks vastav katsetehnika (Anless, Bengtsson, Engqvist, 1995).

Arendades edasi ühes korduses läbiviidavate katsete metoodikat, leidsime, et ühes korduses on võimalik läbi viia ka sordivõrdluskatseid. Kuid ainult ühe korduse olemasolu korral tuleb siin rakendada kahefaktorilist põldkatset – üheks faktoriks on sort, teiseks faktoriks aga mingi kvantitatiivselt muudetav mõjur (väetisenorm, külvisenorm või midagi muud). Näiteks võib korraldada põldkatse 4 erineva teraviljasordiga (A, B, C ja D) 11 lämmastiktootumise tasemel (N_0 , N_{15} , N_{30} , N_{45} , N_{60} , N_{75} , N_{90} , N_{105} , N_{120} , N_{135} ja N_{150}). Variante on sellises põldkatses kokku 44 (4×11).

Variandid võivad olla katses ühes korduses, see tähendab, et iga variant on ainult ühel katselapil, välja arvatud lämmastiväetise nulltase. Iga konkreetne sort on lämmastiväetise nulltasemel kahel katselapil, s.o. kahes korduses. Antud olukord võimaldab regressioonanalüüsil paremini paika panna regressiooni alguspunkti (regressiooni vabaliikme väärtuse). Seega on sellises põldkatses kokku 48 katselappi.

Agronoomiaalastes uuringutes esineb üsna sageli olukordi, millistes selgitatakse eri sortide reageerimist väetise, sealhulgas ka lämmastikväetise toimel. Sisuliselt on tegemist kahefaktoriliste katsetega ning kui kõrvale jätta nüansid, siis tänaseni on põldkatsete struktuur olnud näiteks järgmine: 4 sorti, 4 erinevat väetamise taset (normi) ja 4 kordust. Sellise põldkatse läbiviimiseks on aga vaja vähemalt 64 ($4 \times 4 \times 4$) katselappi. Järelikult, ainult ühe korduse olemasolu korral põldkatses me võidame eksperimentaalse töö mahus, sest katselappe on vähem. Teiseks oluliseks eeliseks on asjaolu, et ühe korduse puhul me võime katsesse võtta tunduvalt rohkem variante (erinevaid väetise norme tunduvalt väiksema sammuga), mis teeb uuritava faktori optimaalse taseme kindlaksmääramise lihtsamaks ja täpsemaks.

Kuidas toimub variantide paigutamine katselappidele? Variandid tuleb katselappidele paigutada randomiseeritult, s.t. juhuslikult ning selleks võiks kasutada niinimetatud *split-plot* meetodit (Little, Hills, 1972). Seda meetodit ei ole Eestis kahefaktoriliste katsete läbiviimisel senini praktiliselt kasutatud ja seetõttu sobiv eestikeelne sõnavaste puudub.

Paigutusskeemil jaotatakse eelkõige katsepõld 12 võrdsesse ossa – vastavalt lämmastikväetise normide arvule. Joonisel on need jaotused piiritletud jämedama joonega. Seejärel seostatakse iga väetuskatselapp kindla väetisenormiga (kas loosimise või juhuslike arvude tabeli abil). Tingimuseks on aga, et N_0 -lapid ei satuks kõrvuti.

Väetuskatselappide siseselt moodustatakse sortidele teist järku katselapid. Selleks tuleb väetuskatselapp antud juhul jaotada neljaks võrdse pindalaga väiksemaks osaks. Sordid paigutatakse selles skeemis samuti juhuslikult. Iga konkreetse väetuskatselapi piires juhuslikku paigutamist korratakse, s. t. sortide järjestus muutub.

Uurimistöö tulemused allutatakse regressioonanalüüsile. Sagedamini kasutatakse selleks ruutvõrrandit:

$$y = a + bx + cx^2, \text{ millises}$$

y = võrrandi põhjal arvutatavad saagid, argumendi funktsioon (nt. saagid, kg/ha),

a = võrrandi vabaliige (nt. saak väetamata mullalt, kg/ha),

b ja c = regressioonikordajad,

x = argument (nt. lämmastiku norm, kg/ha).

N ₉₀ D	N ₆₀ D	N ₁₃₅ A	N ₁₅ B
N ₉₀ B	N ₆₀ A	N ₁₃₅ D	N ₁₅ D
N ₉₀ C	N ₆₀ B	N ₁₃₅ B	N ₁₅ A
N ₉₀ A	N ₆₀ C	N ₁₃₅ C	N ₁₅ C
N ₄₅ D	N ₁₂₀ B	N ₀ D	N ₃₀ C
N ₄₅ C	N ₁₂₀ D	N ₀ A	N ₃₀ B
N ₄₅ B	N ₁₂₀ C	N ₀ C	N ₃₀ A
N ₄₅ A	N ₁₂₀ A	N ₀ B	N ₃₀ D
N ₀ D	N ₁₅₀ A	N ₁₀₅ B	N ₇₅ C
N ₀ B	N ₁₅₀ B	N ₁₀₅ C	N ₇₅ B
N ₀ A	N ₁₅₀ C	N ₁₀₅ A	N ₇₅ D
N ₀ C	N ₁₅₀ D	N ₁₀₅ D	N ₇₅ A

Skeem. Variantide paigutumine kahefaktorilises ühe kordusega põldkatses
Scheme. Layout of an unreplicated field trial with 2 factors

Enne regressioonanalüüsi on soovitatav tegelikud, uurimistöö tulemusena saadud andmed korrigeerida juhusliku vea suhtes. Korrigeerimine toimub rea keskel olevate liikmete jaoks libiseva keskmise arvutamise põhimõttel kolme tegeliku tulemuse baasil. Rea alguses või lõpus olevate liikmete jaoks toimub korrigeerimine kas kahe või kolme tegeliku saagipunkti abil ja vastavalt väljatöötatud meetodikale (Lauk, 1994; 1995).

Korrigeerimine vähendab tunduvalt juhuslikku viga, mistõttu regressioonanalüüsil saadakse tunduvalt suuremad determinatsiooni- ja korrelatsioonikordajate väärtused ning väheneb regressiooni viga *resp.* standardhälve.

Iga konkreetse sordi korrigeeritud andmed allutatakse regressioonanalüüsile, mille tulemusena saadakse 4 regressioonivõrrandit, mis kajastavad sortide reageerimist lämmastikväetise suhtes. Sortide reaktsiooni lämmastikväetise suhtes saab kujutada joonistel graafiliselt. Ühtlasi on võimalik ruutfunktsiooni tuletise abil välja arvutada väetise diferentsiaalefektiivsus, agronoomiliselt optimaalsed või maksimaalsed lämmastikväetise kogused igale konkreetsele sordile eraldi jne.

Sordivõrdlus toimub järgmiselt. Igale konkreetsele, vastava sordi regressioonikõverale (regressioonivõrrandi põhjal arvatud tulemusel) leitakse usalduspiirid valemiga:

$$U_{95\%} = s_y \times t_{\text{tabeliline}(p = 5\%)},$$

$U_{95\%}$ = regressioonikõvera usalduspiirid 95% usutavuse juures (eksimisvõimalus 5%),

s_y = regressiooni standardhälve,

$t_{\text{tabeliline}(p = 5\%)}$ = Studenti teoreetiline kriteerium 5%-lise usalduslääve juures. Võetakse vastavatest tabelitest jääkvariatsiooni vabadusastmete arvu kohaselt.

Sorte võib võrrelda paarikaupa, kusjuures üheks sordiks on alati standardsort. Sortide regressioonikõveraid kujutatakse graafiliselt koos kõvera usalduspiiridega. Kahe sordi saagid erinevad usutavalt, kui kummagi sordi usalduspiirid ei kattu (*resp.*, kui kahele keskmisele arvatud usalduspiirid ei kattu). Võib aga tekkida olukord, et madalamal lämmastiktootumise tasemel sortide saagid usutavalt ei erine (usalduspiirid kattuvad), kuid kõrgemal lämmastiktootumise tasemel regressioonikõverate usalduspiirid ei kattu ning saakides on usutavad erinevused olemas või vastupidi. Sisuliselt selgub ka uuritavate faktorite koosmõju.

Kuid soovi korral võib sordivõrdluseks kasutada ka dispersioonanalüüsi, vaatamata sellele, et katse on ainult ühes korduses. Dispersioonanalüüsi kasutamisel moodustatakse kordused lämmastiktootumise tasemete baasil. Dispersioonanalüüsil on variatsiooni allikateks vastavalt sordid, kordused (sisuliselt küll lämmastikväetise eri tasemed) ja juhuslikkus (viga). Töös toodud näite puhul võib koostada 3 erinevat dispersioonanalüüsi ülesannet:

I ülesanne: sordivõrdlus suhteliselt madalal lämmastiktootumise tasemel. Kordused sordivõrdluseks moodustatakse järgmiste lämmastiktootumise tasemete baasil: 1. kordus = N_0 saak sortidel, 2. kordus = N_0 saak, 3. kordus = N_{15} saak ja 4. kordus = N_{30} saak.

II ülesanne: sordivõrdlus keskmisel lämmastiktootumise tasemel. Kordused sordivõrdluseks moodustatakse vastavalt: 1. kordus = N_{45} saak sortidel, 2. kordus = N_{60} saak, 3. kordus = N_{75} saak ja 4. kordus = N_{90} saak.

III ülesanne: sordivõrdlus suhteliselt kõrgel lämmastiktootumise tasemel. Kordused sordivõrdluseks moodustatakse vastavalt: 1. kordus = N_{105} saak sortidel, 2. kordus = N_{120} saak, 3. kordus = N_{135} saak ja 4. kordus = N_{150} saak.

Kohustuslik ei ole moodustada tingimata 4 dispersioonanalüüsi ülesannet, vaid neid võib koostada ka vähem ja just meid huvitava uuritava faktori tasemete (väetisenormide) piirkonnas. Viimasel juhul võib dispersioonanalüüsi ülesande koostada suurema arvu kordustega.

Kirjeldatud meetodit on EPMÜ taimekasvatuse instituudis autori soovitusel katsete läbiviimisel ja andmete töötlemisel kasutatud, kuid antud juhul ei saa neid näiteks tuua, sest originaalmaterjal ei kuulu autorile.

Kirjandus

- Alness, K., Bengtsson, A., Engqvist, A. An application and grading method of field trials with linear change of dose. – Proceedings of the Fourth Regional Conference on Mechanisation of field experiments (IAMFE/BALTIC '95). Kaunas/Dotnuva, Lithuania, August 8-10, 1995. Uppsala, Sweden, July, p. 86...92, 1995.
- Lauk, E. Uut põldkatsete meetodikas ja andmetöötluses. – Eesti Põllumajandusülikooli teadustööde kogumik, nr. 178. Agronoomia. – Tartu, Eesti Põllumajandusülikool, lk. 147...152, 1994.
- Lauk, E. Regression analysis: A good method for analysing the field experiments data. – Proceedings of the Fourth Regional Conference on mechanisation of field experiments (IAMFE/BALTIC '95). Kaunas/Dotnuva, Lithuania, August 8-10, 1995. Uppsala, Sweden, July 1995, p. 35...41, 1995.
- Little, T. M., Hills, F. J. Statistical methods in agricultural research. – Berkeley, California, 1972. – 242 p.

Regression Analysis in Variety Tests

E. Lauk

Summary

This article gives a survey of the possibilities for using regression analysis in processing variety test data. Regression analysis is an effective method, but it presumes the use of a somewhat different structure for planning the experiment. The field experiment must have two components (factors) – one is the variety, the other is a quantitative change (norm of fertilizer, seed norm, or some other factor). If we use regression analysis, we can experiment with one replication, increasing significantly the

number of levels for the quantitative changes. The total number of experimental plots and the amount of experimental work to be done diminishes.

The establishment of the experimental plots must be carried through in one block with a split-plot design. The first step is to randomize the different levels of the quantitative changes on the main plots. On the basis of the main plots the subplots are formed, and randomization is used to place the varieties on those plots.

The result of the regression analysis will give regression equations for every variety, that illustrate the reactions of the varieties towards the quantitative changes. The reactions towards the quantitative changes could be shown in figures. For the regression curves of the varieties the confidence limits will be calculated, using the standard deviations (s_y) and the corresponding theoretical criteria (the t-test for example). The confidence limits will also be shown in the figures. A significant difference exists between the two varieties if the confidence limits of the two varieties do not coincide. There could also be a situation where the varieties are not different at the lower levels of quantitative changes, but at higher levels the confidence limits of the regression curves do not coincide, and there is a significant difference in the yields, or *vice versa*.

It is possible to use variance analysis for the comparison of the varieties, where the replicas would be formed by different levels of quantitative changes. The source of the variance analysis is formed by the varieties, the replicas (essentially the different levels of quantitative changes), and randomization (error).